A syntax for semantics in P-Lingua

Ignacio P´erez-Hurtado, David Orellana-Martín, Agust´ın Riscos-N´u˜nez, and Mario J. P´erez-Jim´enez

Research Group on Natural Computing
Department of Computer Science and Artificial Intelligence
Universidad de Sevilla, Spain
{perezh,dorellana,ariscoan,marper}@us.es

Summary. P-Lingua is a software framework for Membrane Computing, it includes a programming language, also called P-Lingua, for writing P system definitions using a syntax close to standard scientific notation. The first line of a P-Lingua file is an unique identifier defining the variant or model of P system to be used, i.e, the semantics of the P system. Software tools based on P-Lingua use this identifier to select a simulation algorithm implementing the corresponding derivation mode. Derivation modes define how to obtain a configuration $C_{t+1}$ from a configuration $C_t$. This information is usually hard-coded in the simulation algorithm.

The P system model also defines what types or rules can be used, the P-Lingua compiler uses the identifier to select an specific parser for the file. In this case, a set of parsers is codified within the compiler tool. One for each unique identifier.

P-Lingua has grown during the last 12 years, including more and more P system models. From a software engineering point of view, this approximation implies a continuous development of the framework, leading to a monolithic software which is hard to debug and maintain.

In this paper, we propose a new software approximation for the framework, including a new syntax for defining rule patterns and derivation modes. The P-Lingua users can now define custom P system models instead of hard-coding them in the software. This approximation leads to a more flexible solution which is easier to maintain and debug. Moreover, users could define and play with new/experimental P system models.

1 Introduction

Membrane computing is an unconventional model of computation within natural computing that was introduced in 1998 by Gh. P˘ aun [17]. The computational devices in membrane computing, also known as P systems, are non-deterministic theoretical machines inspired on the biochemical processes that take place inside the compartments of living cells.

Several kinds of P systems coexist, each of them having different syntactic ingredients, such as different alphabets and structures. The two most studied are
cell-like membrane systems, characterized by their rooted tree structure, where membranes act as filters that let certain elements to pass through them [17], and membrane systems structured as directed graphs, representing the communication between cells within a tissue of a living being, called tissue-like P systems [9] or between neurons in a brain, called spiking neural P systems [7]. The interchange of objects between the different compartments is defined by the rules of the system, that together with the corresponding semantics, mark the functioning of the system.

A configuration of a P system is defined by the structure of the compartments at a certain moment, and the elements (being usually objects, although other kinds of elements can be considered, as strings, catalysts [17] and anti-matter [14], among others) contained in each compartment, as well as other characteristics from specific types of P systems, providing a snapshot of the system at an instant $t$. By using the rules specified in a model, we can make its objects change, both evolving and moving between the different compartments (membranes in the case of cell-like P systems and cells in the case of tissue-like P systems).

On the one hand, in P systems with active membranes [19], both objects and membranes change through the application of evolution, communication, division, separation, creation and dissolution. In this framework, membranes can have a polarization associated to each membrane. On the other hand, in tissue P systems [9], symport/antiport rules are devoted to make objects move from a cell to another cell or to the environment (a special compartment where there exist an arbitrary number of objects of an alphabet defined a priori), while division and separation rules allow an exponential growth in linear time.

We say that a configuration $C_t$ yields to a configuration $C_{t+1}$ if, by applying the rules specified in the model according to its semantics, we can obtain $C_{t+1}$ from $C_t$. Semantics rules the behavior of the system, determining which rules can be applied and how they affect the system according to a global clock. A computation of a P system is a (finite or infinite) sequence of instantaneous configurations.

We consider a family (or model) of P systems as the definition of a type of P system, that is, its syntax and semantics. According to the specification of a particular family of P systems, we consider a (specific) model as the definition of an individual P system, that is, its working alphabet, initial membrane structure with initial multisets of objects and the set of rewriting rules with another characteristics of the correspondent family. By the definition of the family, we can interpret the structure and behavior of a specific model within that family.

Membrane computing is a very flexible framework where different types of devices can be outlined. In fact, the intersection between Membrane Computing and other fields, such as engineering [20], biology [23] and ecology [2], as well as a long list of other scientific lines [5, 13, 24], has generated necessities that could only be filled by the creation of new kinds of P systems, expanding the scope of researchers in this area. For an exhaustive explanation of the different types of P systems, we refer the reader to [18] and [16].
In this work, we have reinvented the P-Lingua framework [6, 25] to include semantic features concerning to the models.

The paper is structured as follows: In the next section, some preliminaries concepts about P-Lingua are introduced. In Section 3, we propose an extension for the P-Lingua language to directly define model constraints in the own P-Lingua files, providing a more flexible and experimental framework. The next Section is devoted to the new GNU GPLv3 software tool to compile the input P-Lingua files. In Section 5 some examples of the new P-Lingua extension are introduced. Finally, some conclusions and future work are drawn.

2 Preliminaries

P-Lingua [6, 25] is a software framework that includes a definition language for P systems (also called P-Lingua) and a GNU GPLv3 Java library (pLinguaCore) that is able to parse P-Lingua files and simulate computations. The library contains three main components:

- A parser for reading input files in P-Lingua format and checking syntactic and semantic constraints related to predefined models. In order to achieve this, the first line of a P-Lingua file should include a P system model declaration by using an unique identifier. There are several P system models that can be used, each one with its own identifier, such as transition, membrane, division, tissue systems, and probabilistic. The analysis of semantic ingredients, such as rule patterns, is hard-coded for each model. Several versions of pLinguaCore [6, 8, 10, 21] have been launched to cover different types of models.
- For each type of model, the pLinguaCore library includes one or more built-in simulators, each one implementing a different simulation algorithm. For instance, Population Dynamic P systems [1] (probabilistic identifier in P-Lingua) can be simulated within the library by applying three different algorithms: BBB, DNDP, and DCBA [3, 11]. Remarkable software projects such as McCoSim (Membrane Computing Simulator) [27, 22] use the simulators integrated in the library to perform P system computations and generate relevant information as result for custom applications.
- Alternatively, the pLinguaCore library is able to transform the input P-Lingua files to other formats such as XML or binary format in order to feed external simulators. The generated files for the given P systems are free of syntactic/semantic errors since the transformation is done after the parser analysis. Several external simulators use this feature, for example, the PMCGPU project (Parallel simulators for membrane computing on GPU) [12, 26] uses definitions generated by pLinguaCore in order to provide the input of CUDA GPU simulators.

The P-Lingua language is currently a standard widely used for the scientific community since the syntax is modular, parametric and close to the common
scientific notation. The description of the language can be found in the references [6, 8, 10, 21, 25]. As an example, the definition of a basic transition P system follows:

```model<transition>
def main()
{
@mu = [[[1]3 [1]4]2]1;
@ms(3) = a,f;
[a --> a,bp]3;
[a --> bp,@d]3;
[f --> f*2]3;
[bp --> b]2;
(1) [f*2 --> f ]2;
(2) [f --> a,@d]2;
}
```

In the example, a module `main` is defined including an initial membrane structure `[ [ ] 3 [ ] 4 ] 2 ] 1`, an initial multiset for the membrane labelled 3, and a set of seven multiset rewriting rules. The special symbol `@d` is used to specify dissolution. The last two rules include priorities as integer numbers in parenthesis at the beginning of the left-hand side of the rules. More complex examples can be found in the P-Lingua web [25].

3 An extension of P-Lingua for semantic features

As explained above, the analysis of semantic ingredients belonging to P systems is hard-coded in the pLinguaCore library, i.e., the only way to define new types of models is by implementing code inside the library. In this section, we propose an extension for the P-Lingua language to directly define model constraints in the own P-Lingua files, providing a more flexible and experimental framework. Two types of semantic constraints can be defined with this extension: rule patterns and derivation modes.

3.1 Rule patterns

The P-Lingua parser is able to recognize rules with the next general syntax:

\[
P \rightarrow \alpha \mu[v_1,v_1,1]^{\alpha_1,1} \cdots [v_1,m_1]^{\alpha_1,m_1}_{h_1,1} \cdots [v_n,v_n,1]^{\alpha_n,1} \cdots [v_n,m_n]^{\alpha_n,m_n}_{h_n,1} \cdots [v_n,m_n]^{\alpha_n,m_n}_{h_n} \]

where:

- $p$ is a priority related to the rule given by a natural number, where a lower number means a higher rule priority.
- $q$ is a probability related to the rule given by a real number in $[0, 1]$.
- $\alpha_i, \alpha_{i,j}, 1 \leq i \leq n, 1 \leq j \leq m_i$ and $\beta_i, \beta_{i,j}, 1 \leq i \leq s, 1 \leq j \leq r_i$ are electrical charges.
- $h_i, h_{i,j}, 1 \leq i \leq n, 1 \leq j \leq m_i$ and $g_i, g_{i,j}, 1 \leq i \leq s, 1 \leq j \leq r_i$ are membrane labels.
- $u, v_i, v_{i,j}, 1 \leq i \leq n, 1 \leq j \leq m_i$ and $w_i, w_{i,j}, 1 \leq i \leq s, 1 \leq j \leq r_i$ are multisets of objects.

Next, there is a list of P-Lingua rule examples matching the general rule syntax:

- $a, b \ [ d, e*2 ] \ 'h$ --> $[ f, g ] \ 'h ; q$; where $q$ is the probability of the rule.
- $(p) \ [a] \ 'h$ --> $[b] \ 'h ;$ where $p$ is the priority of the rule.
- $[a \ --> \ b] \ 'h ;$ the left-hand side and right-hand side of evolution rules can be collapsed.
- $+[a] \ 'h$ --> $+[b] \ 'h \ -[c] \ 'h ;$ a division rule using electrical charges.
- $[a] \ 'h$ --> $[b] \ 'h ;$ a dissolution rule.
- $a[ ] \ 'h$ --> $[b] \ 'h ;$ a send-in rule.
- $[a] \ 'h$ --> $b[ ] \ 'h ;$ a send-out rule.
- $[a \ --> \ #] \ 'h ;$ the symbol $#$ can be optionally used as empty multiset.
- $[a] \ '1 \ <--> \ [b] \ '0 ;$ a symport/antiport rule in the tissue-like framework.

The syntax of the general rule is very permissive, and so different parsers for different models have been developed in order to restrict the rules used in each one. In order to provide the researcher a more flexible framework, not having to depend on the implementation itself but acquiring the capacity of restricting the model by himself, we introduce the next syntax in P-Lingua for rule pattern matching:

```plaintext
!rule-type-id
{
 pattern1
 pattern2
 ...
 patternN
}
```

where `rule-type-identifier` is an unique name for the type of rule that is going to be defined and `pattern1`, `pattern2`, ..., `patternN` are rule patterns following the same syntax than common rules in P-Lingua where anonymous variables beginning
with ? can be optionally used instead of probabilities, charges and priorities. In the patterns, the symbols beginning with a, b or c always mean single objects and symbols beginning with u, v and w always mean multisets of objects. In Section 5, are given several examples of rule patterns in P-Lingua for different types of cell-like and tissue-like models.

3.2 Derivation modes

From an informal point of view, we can see a derivation mode as the way a step of a P system is performed. As a part of semantics, it rules the exact application of rules of the system, deciding when rules can be applied or not when they are applicable. An extensive study of derivation modes can be found in [4]. In order to make the work self-content, we give a minimal definition of a derivation mode.

A derivation mode \( \vartheta \) is defined as a function that selects different multisets of rules “really applicable” to a configuration \( C_t \) of a P system depending on a specification. For this purpose, let \( H \) be a P system with \( R \) as its set of rules, \( R \) a multiset of compatible rules applicable to a P system at configuration \( C_t \), and let \( R \) be the set of all multisets applicable to a P system at configuration \( C_t \).

In this extension of P-Lingua we provide two main derivation modes:

- **Maximally parallel derivation mode** \((\text{max})\): It is the default mode for P systems. In this mode, we only take multisets from \( R \) that are not extensible, that is:
  \[
  R' = \{ R \mid R \in R \land \nexists R' \in R : R \subseteq R' \}.
  \]
  The multiset of rules finally applied to \( C_t \) is selected non-deterministically from \( R' \).

- **Bounded-by-rule parallel derivation mode** \((\text{bound}_{B_1,\ldots,B_r})\): Let \( \{a, b, \ldots\} \) be the set of different types of rules present in a P system. \( B_i \) can be of the following forms:
  - \( B_i = j, j \in \{a, b, \ldots\} \);
  - \( B_i = \beta_n(B_{1i}, \ldots, B_{ri}) \), being \( n \in \mathbb{N} \), and for each \( B_j = \beta_m(B_{1j}, \ldots, B_{mj}) \), \( j \in \{1, \ldots, r_i\}, m_j \leq n \);
  - As a restriction, a label for a type of rule cannot appear more than once in the whole definition of the derivation mode.

We say that \( n \) is the bound of \( B_i = \beta_n \). We say that a type of rule \( (j) \) is in the context of \( B_i \) if:
  - There exists \( B_i = \beta_n(j) \) (we call \( B_i \) its immediate context); and
  - There exists \( B_j = \beta_n(B_{1j}, \ldots, B_{mj}) \) such that \( B_j \) is a context of the type of rule \( (j) \).

This mode is defined recursively, and we can understand the applicability of the rules in a defined bounded-by-rule parallel derivation mode in the following sense:
– In a context $\beta_n(B_1, \ldots, B_r)$, the number of rules that can be applied in parallel in a P system in a configuration $C_t$ is $n$; and
– In a bounded-by-rule parallel derivation mode $bound(B_1, \ldots, B_r)$, if $B_i = j (j \in \{a, b, \ldots\})$, being $1 \leq i \leq r$, then rules of type $j$ can be applied in a maximal way.

With this mode, we can define the classical mode used in P systems with active membranes, that is, evolution rules $(a)$ can be applied in a maximal parallel mode, while the other types of rules (send-in communication rules $(b)$, send-out communication rules $(c)$, dissolution rules $(d)$, division rules for elementary $(e)$ and non-elementary $(f)$ membranes) can be applied at most once per membrane at each computation step. It would be defined as $bound_{a,\beta}$. If $R_j$ is the set of rules from $R$ of the type $j$, we formally define the bounded-by-rule maximally parallel mode by

$$R' = \{ R \mid R \in R \land |\{r \mid r \in R, r \in R_j\}| \leq n \text{ for all } j \text{ in the context of } B_i = \beta_n \land \beta R' \in R : R \subseteq R' \}$$

Thus, a model type can be defined in P-Lingua by aggregating the allowed rule patterns and its corresponding derivation modes, the syntax is as follows:

```
@model(id) = rule-type-id1,..., rule-type-idN;
```

where $id$ is an unique identifier for the model and $rule-type-id1$, ..., $rule-type-idN$ are unique identifiers for the corresponding allowed rule patterns. By default all rules behave in maximally parallel derivation mode, but rules can be grouped in sets to behave in bounded parallel derivation mode as follows:

```
@model(id) = @bound{rule-type-id,..., rule-type-idN};
```

where $bound$ is a natural number defining the maximum number of rules in the group that can be applied to a given configuration. In Section 5, several examples of model definitions in P-Lingua are given.

### 4 Command-line tools

A set of two GNU GPLv3 command-line tools called `plingua` and `psim` have been implemented in C++ language with Flex [28] and Bison [29]. The source code including examples and instructions can be downloaded from https://github.com/RGNC/plingua.

The tool provides three main functionalities:
• **Parsing P-Lingua files** while printing the syntactic and semantic errors to the standard error output. In this sense, the tool acts as a conventional compiler, showing the name of the file, as well as the number of the line and column for each error with a short description. The analysis of semantic errors is done by using the rule patterns and derivation modes defined in the own P-Lingua files. Several files can be compiled together like conventional programs, furthermore standard *makefiles* can be also used. The developer can decide to write the rule patterns and derivation modes in a set of files and reuse them in several projects. More explanations can be found in the website.

• **Generating JSON/XML/Binary files.** The tool is able to translate the definitions contained in P-Lingua files to standard formats such as JSON, XML and Binary (compressed bit-level file format) for compatibility with third-party simulators. The conversion is done after parsing the input files, thus the output files are free of syntactic/semantic errors and the third-party applications do not have to check them. Several P-Lingua files can be combined together in one output file, including also the selected derivation modes.

• **Simulation of P system computations.**

Detailed information about how to use these tools, including examples, can be found in the website of the project [https://github.com/RGNC/plingua](https://github.com/RGNC/plingua).

5 Examples

Next, there are some examples of rule patterns and definition of derivation modes for several common P system models. Please, see the website of the project for more information.

5.1 Transition P systems

```plaintext
 transition_evolution /* Limited to rules with 3 inner membranes */ { 
   [a --> v]';h;
   [a --> v, 0d]';h;
   (??) [a --> v]';h;
   (??) [a --> v, 0d]';h;
   [a [ ]'h1 --> v [w]']h1';h;
   (??) [a [ ]'h1 --> v [w]']h1';h;
   (??) [a [ ]'h1 --> v [w]']h1';h;
   (??) [a [ ]'h1 --> v [w]']h1';h;
   (??) [a [ ]'h1 --> v [w]']h1';h;
   [a [ ]'h1 [ ]'h2 --> v [w1]'h1 [w2]'h2']';h;
   (??) [a [ ]'h1 [ ]'h2 --> v [w1]'h1 [w2]'h2']';h;
   (??) [a [ ]'h1 [ ]'h2 --> v [w1]'h1 [w2]'h2']';h;
   (??) [a [ ]'h1 [ ]'h2 --> v [w1]'h1 [w2]'h2']';h;
   [a [ ]'h1 [ ]'h2 [ ]'h3 --> v [w1]'h1 [w2]'h2 [w3]'h3']';h;
   (??) [a [ ]'h1 [ ]'h2 [ ]'h3 --> v [w1]'h1 [w2]'h2 [w3]'h3']';h;
   (??) [a [ ]'h1 [ ]'h2 [ ]'h3 --> v [w1]'h1 [w2]'h2 [w3]'h3']';h;
}
```

@model(transition) = transition_evolution;

5.2 Active membranes with division rules

```
!dam_evolution
{
  ?[a -> v]’h;
  ?[a -> ]’h;
}

!dam_send_in
{
  a ?[ ]’h -> ?[b]’h;
}

!dam_send_out
{
  ?[a]’h -> b ?[ ]’h;
}

!dam_dissolution
{
  ?[a]’h -> b;
  ?[a]’h -> ;
}

!dam_division
{
  ?[a]’h -> ?[ ]’h ?[ ]’h;
  ?[a]’h -> ?[b]’h ?[ ]’h;
  ?[a]’h -> ?[ ]’h ?[b]’h;
  ?[a]’h -> ?[b]’h ?[c]’h;
}

@model(membrane_division) =
  dam_evolution, @1{dam_send_in, dam_send_out, dam_dissolution, dam_division};
```

5.3 Tissue-like P systems with communication and cell division

```
!tissue_communication
{
  [u]’h1 <-> [v]’h2;
}

!tissue_division
{
  [a]’h -> [ ]’h [ ]’h;
  [a]’h -> [b]’h [ ]’h;
  [a]’h -> [ ]’h [b]’h;
  [a]’h -> [b]’h [c]’h;
}

@model(tissue_division) =
  tissue_communication, @1(tissue_division);
```

5.4 Population Dynamics P Systems

```
!pdp_evolution
{
  u1 ?[v1]’h -> u2 ?[v2]’h :: ?;
}

!pdp_environment_communication
{
  [a]’e1 [ ]’e2’h -> [[ ]’e1 [b]’e2]’h :: ?;
```
Conclusions and future work

This paper reinvents P-Lingua moving forward to a more flexible tool which is easier to maintain and debug. The goal is twofold: On the one hand, it pretends to be a good assistant for researchers while verifying their designs, even working with experimental models. On the other hand, more general simulators can be developed, covering a large set of P system variants by reading and simulating the custom derivation modes.

Several lines are open for future work. From the point of view of the language, the semantic ingredients that can be written in P-Lingua should be studied in order to cover more types of models. For instance, defining bounds for the multiplicities of objects in different compartments, such as the environment in tissue-like P systems, where the multiplicity of objects can be infinite. On the other hand, custom directives could be included in P-Lingua files and translated to C preprocessor directives for the simulator. For example, if the design is confluent, a directive could be written to optimize the simulation time, since it is not necessary to simulate the non-determinism by using random numbers.

From the point of view of the simulation tools, we are interested about the integration with CUDA, OpenMP, FPGA and POSIX threads, optimizing the use of parallel architectures.

Acknowledgements

The authors acknowledge the support of the research project TIN2017-89842-P, co-financed by Ministerio de Economía, Industria y Competitividad (MINECO) of Spain, through the Agencia Estatal de Investigación (AEI), and by Fondo Europeo de Desarrollo Regional (FEDER) of the European Union.

References


28. The Flex web page: [https://github.com/westes/flexl](https://github.com/westes/flexl)
29. The Bison web page: [https://www.gnu.org/software/bison/](https://www.gnu.org/software/bison/)